**November 8, 2011**

**Tuuli Lappalainen, tuuli.lappalainen@unige.ch**

**University of Geneva**

**Geuvadis RNA sequencing guidelines**

**Summary of the most critical points**

* We have 465 RNA samples in total going into sequencing.
* All the partners will do the mRNA and miRNA library prep of the samples that have been allocated to them and sequence them in full coverage. In addition to this, Geneva will do mRNA library prep also of the samples that are sequenced in other labs, and low-coverage sequencing of these libraries.
* Barcelona will be the first lab to start miRNA library preps, and will give further specifications of the protocol later
* All sequencing is to be done with HiSeq
* The library preparation has to be done in a random order within each lab
* mRNAseq: 75 bp PE sequencing to at least 10M mapped and properly paired read pairs. Library prep kit TruSeq RNA Sample Prep Kit v2, sequencing kit v3
* miRNAseq: 36 SE sequencing to 3-6M mapped reads. Library prep kit TruSeq Sm RNA Sample Prep, sequencing kit v3
* Any modifications in the protocol during the experiment should be avoided – but if you change anything, please keep track and report
* The labs should aim at submitting fastq files to EBI by **February 15**, 2012.

If something is unclear, please don’t hesitate to ask: [tuuli.lappalainen@unige.ch](mailto:tuuli.lappalainen@unige.ch), +41-22-3795550

**Table 1: Sample allocation to Geuvadis partners**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Lab #** | **mRNAseq lib prep** | **mRNAseq sequencing** | **miRNAseq lib prep & seq** |
| UNIGE (Geneva) | 1 | 465 | 116 full coverage,  354 half coverage | 116 |
| CRG (Barcelona) | 2 | 96 | 96 | 96 |
| MPIMG (Berlin) | 3 | 72 | 72 | 72 |
| ICMB (Kiel) | 4 | 72 | 72 | 72 |
| HMGU (Munich) | 7 | 48 | 48 | 48 |
| UU (Uppsala) | 5 | 48 | 48 | 48 |
| LUMC (Leiden) | 6 | 48 | 48 | 48 |

Counts include 5 duplicate samples that will be sequenced in full coverage every lab

**Table 2: RNA samples by population**

|  |  |  |
| --- | --- | --- |
|  | **Geuvadis** | **Geuvadis + 1KG Phase1** |
| TSI | 93 | 92 |
| GBR | 96 | 86 |
| FIN | 95 | 89 |
| CEU | 92 | 79 |
| YRI | 89 | 77 |
| TOTAL | 465 | 423 |

**RNA samples**

* Total RNA was extracted with Trizol in Geneva in October
* Qbit and total RNA Bioanalyzer has been done in Geneva to estimate RNA quality and quantity
* The laboratories will receive a minimum of 4 ug of RNA of the samples allocated to them. The samples will be shipped from Geneva on November 7, 2011
* All laboratories should do total RNA Bioanalyzer for 10-20% of the RNA samples before library prep
* Library preps should be done in a random order in every lab. The sample information sheets are already randomized – do **not** sort them

**mRNAseq:**

* All the partners will do the library prep of the samples that were allocated to them and sequence their libraries from that prep. No libraries will be shared.
* In addition to this, Geneva will do library prep of all the samples that are sequenced in other labs, and sequence these libraries in 1/3 – 1/2 coverage
* These allocations include 5 samples that will be done in full coverage in every lab
* Sequencing should be done with HiSeq
* 75 bp paired-end sequencing with fragment size of 280 bp. If you cannot do 75bp, you can sequence 100bp PE and trim the reads.
* Coverage: minimum 10M mapped and properly paired read pairs. This number should be estimated from reads trimmed to 50bp and paired-end mapped with bwa (or an equivalent mapper) to the reference genome (hg19), without filtering for mapping quality.
* Library prep kit: TruSeq RNA Sample Prep Kit **v2**, the high-throughput protocol
* Cluster generation kit: TruSeq PE Cluster Kit **v3**
* Sequencing kit: TruSeq SBS Kit **v3**
* The partners can choose freely how to pool the samples to get the desired coverage
* Bioanalyzer should be run for all library preps before sequencing

**miRNAseq**

* All the partners will do the library prep and sequence that in full coverage (3-6M mapped reads) of the samples that were allocated to them
* These allocations include 5 samples that will be done in full coverage every lab
* Barcelona will start miRNAseq first to estimate the quality of the RNA, and give further specifications of the method if necessary – others should wait for information from them before starting the miRNA library preps
* Sequencing should be done with HiSeq
* 36 bp single-end sequencing, 3M mapping reads per sample
* Library prep kit: TruSeq Sm RNA Sample Prep
* Cluster generation kit: TruSeq PE Cluster Kit **v3**
* Sequencing kit: TruSeq SBS Kit **v3**

**Sample processing information**

* All the labs need to fill in information of sample processing for both mRNA and miRNAseq in order to enable control of batch effects. These information include
  + Provided by Geneva: RNA extraction batch, RNA Bioanalyzer quality assessment (0/0.5/1), RIN, RNA concentration, sequencing lab number
  + Provided by each lab: RNA quantity in library prep (ng), Library prep date, Library prep batch ID, Library prep technician, Library prep kit batch number, Primer index, Method used in library quantification, Library concentration, Library size bp, Library quantity used in sequencing, Cluster kit batch number, Sequencing kit batch number, Seq run date, Flow cell, Lane, HiSeqID, Raw cluster density, Passed cluster density, HCS version, RTA version
  + Details of these will be given with the sample information sheets

**Data processing**

* Each lab needs to submit one demultiplexed fastq file per sample per mRNAper and miRNAseq, produced by CASAVA 1.8 allowing one mismatch in the index
* The fastq files for should be named as: [SAMPLE\_ID].[SeqLabNumber].[mRNA or miRNA: M/MI]\_[SequencingDate:YYMMDD]\_[Lane]\_[which read: 1 / 2 (always 1 for miRNA)].fastq.gz . Example: NA12345.1.MI\_111215\_3\_1.fastq.gz would be sample NA12345 miRNA-sequenced in Geneva on 2011-Dec-15 on lane 3.
* Instructions for data submission will follow

**Data submission deadline**

* All fastq files and sample information should to be submitted to EBI ENA as soon as they are finished. **The deadline is February 15, 2011**
* If there are unexpected delays in your data production, inform the other partners (especially Geneva) as soon as possible